The Effect of Noise on the Number of Extreme Points

Dominique Attali, Olivier Devillers, Xavier Goaoc

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

The Effect of Noise on the Number of Extreme Points

Dominique Attali — Olivier Devillers — Xavier Goaoc

N° 7134

Décembre 2009

ISSN 0249-6399 ISRN INRIA/RR--7134--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
SOPHIA ANTIPOLIS - MÉDITERRANÉE

The Effect of Noise on the Number of Extreme Points

Dominique Attali* , Olivier Devillers, Xavier Goaoc

Thème : Algorithmique, calcul certifié et cryptographie
Équipes-Projets Geometrica et Vegas

Rapport de recherche n° 7134 — Décembre 2009 — 24 pages

Abstract: Assume that Y is a noisy version of a point set X in convex position. How many vertices does the convex hull of Y have, that is, what is the number of extreme points of Y?

We consider the case where X is an (ϵ, κ)-sample of a sphere in $\mathbb{R}^d$ and the noise is random and uniform: Y is obtained by replacing each point $x \in X$ by a point chosen uniformly at random in some region $\mathfrak{P}(x)$ of size δ around x. We give upper and lower bounds on the expected number of extreme points in Y when $\mathfrak{P}(x)$ is a ball (in arbitrary dimension) or an axis-parallel square (in the plane). Our bounds depend on the size n of X and δ, and are tight up to a polylogarithmic factor. These results naturally extend in various directions (more general point sets, other regions $\mathfrak{P}(x)\dots$).

We also present experimental results, showing that our bounds for random noise provide good estimators of the behavior of snap-rounding, that is when Y is obtained by rounding each point of X to the nearest point on a grid of step δ.

Key-words: Complexity, Convex hull, Noisy data

This work is partially supported by ANR Project *Triangles* ANR-07-BLAN-0319 and by ANR Project *GIGA* ANR-09-BLAN-0331-01.

* Gipsa-lab, CNRS UMR 5216, Grenoble, France. Dominique.Attali@gipsa-lab.inpg.fr

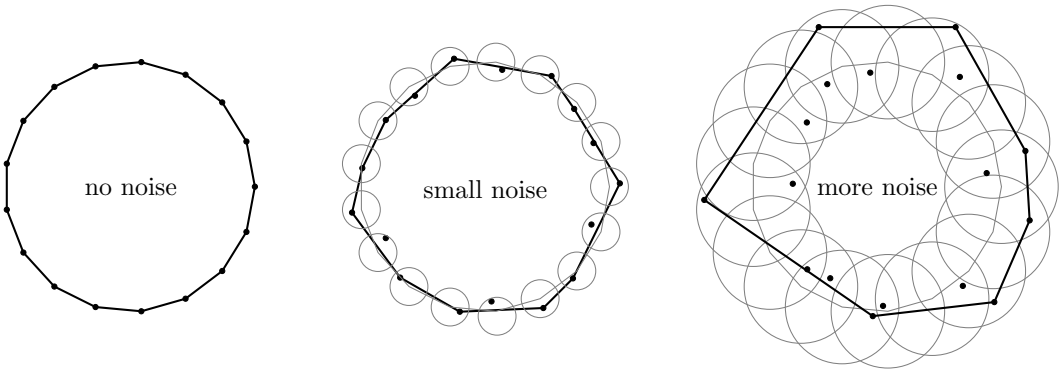
Centre de recherche INRIA Sophia Antipolis – Méditerranée
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Influence du bruit sur le nombre de points extrêmes

Résumé : Si Y est une version bruitée d'un ensemble de points X en position convexe, quelle est la taille de l'enveloppe convexe de Y, c'est à dire quel est le nombre de points extrêmes de Y ?

Nous considérons le cas où X est un (ϵ, κ)-échantillon de la sphère de $\mathbb{R}^d$ et où le bruit est aléatoire et uniforme : Y est obtenu en remplaçant chaque point $x \in X$ par un point choisi aléatoirement dans une région $\mathfrak{P}(x)$ de taille δ au voisinage de x. Nous donnons des bornes inférieures et supérieures sur l'espérance du nombre de points extrêmes de Y lorsque $\mathfrak{P}(x)$ est une boule (en toute dimension) ou un carré (dans le plan). Nos bornes dépendent de n la taille de X et de δ. L'écart entre les bornes supérieurs et inférieures est réduit à un facteur polylogarithmique. Ces résultats peuvent s'étendre dans diverses directions (ensemble de points plus généraux, autre formes pour $\mathfrak{P}(x)\ldots$).

Nous présentons également des résultats expérimentaux montrant que nos bornes pour un bruit aléatoire donnent une bonne estimation du comportement de l'arrondi de points sur une grille. Dans ce cas Y est obtenu en arrondissant chaque point de X au plus proche sommet d'une grille de pas δ.



Mots-clés : Complexité, Enveloppe convexe, Données bruitées

1 Introduction

Understanding the complexity of a data structure or an algorithm is a central problem in computer science. Traditionally, this complexity is measured by the *worst-case* scenario, which gives a bound on how bad things can get. In computational geometry, the input is usually given as a sequence of n reals (e.g. the coordinates of a set of points) and the worst-case complexity of a structure is analyzed as a function of n. In many applications, however, the data is given only at some fixed precision and is subject to measurement noise. This may drastically limit the relevance of lower bounds achieved by unstable constructions over the reals, as even mild restrictions on the complexity of the input representation can forbid certain configurations from arising. For instance, there are order types on n points that can only be realized by coordinates with exponential bit-complexity [14]. As a result, the worst-case configurations may seldom or never occur, and the worst-case complexity may be too general to reflect the practical behavior of the structure or algorithm.

When the worst-case complexity appears too pessimistic, a standard approach is to restrict the input model to make it more *realistic* [7]. For instance, the discrepancy between the worst-case quadratic complexity of the Delaunay triangulation of points in $\mathbb{R}^3$ and the near-linear behavior observed in the context of mesh generation [2] or surface reconstruction [3] can be accounted for by restricting the input to "reasonable" samples of "reasonable" surfaces [1, 8, 9, 10, 12, 13]. There are, however, contexts in which realistic input models remain elusive; for example, in spite of much efforts there is still no satisfying model of "plausible" computer graphics scenes.

An alternative to the application-specific approach of realistic input models is to refine the complexity analysis to take into account more than just the maximum of the complexity function. The *smoothed complexity* model of Teng and Spielman [19] is a step in this direction. The smoothed complexity is defined, informally, as the maximum over the inputs of the expected complexity over small perturbations of that input. Intuitively, this "local averaging" mechanism disposes of configurations that vanish under small perturbation, and may thus model more accurately behaviors on noisy data. Teng and Spielman proved that the simplex algorithm has polynomial smoothed complexity, explaining the apparent difficulty to construct exponential lower bounds and the effectiveness of this algorithm on practical input. Since this seminal paper, smoothed analysis has been applied to a variety of problems [20], including, in computational geometry, the complexity of visibility maps in a terrain [6] and the number of extreme points (i.e. the vertices of the convex hull) [5]. In this paper, we revisit the smoothed complexity of the number of extreme points.

Problem statement and previous work. Let X be a set of n points in $\mathbb{R}^d$. A *perturbation* of X is a point set Y obtained by replacing each point x in X by some point chosen randomly according to some probability distribution $\mathfrak{P}(x)$ centered in x; when $\mathfrak{P}(x)$ is the uniform distribution on the L^p ball of radius δ centered in x, we call the perturbation the *uniform L^p perturbation of amplitude δ*. Damerow and Sohler [5] showed that the expected number of extreme points in a uniform L^∞ perturbation of amplitude δ of a set of n points is $O\left(\left(\frac{n \log n}{\delta}\right)^{\frac{d}{d+1}}\right)$. Their approach is to count the points that are

not necessarily extreme, but *maximal* (i.e. are the corner of some empty axis-parallel octant). Note that the expected number of maxima may be much larger than the expected number of extreme points, e.g. for random samples of a disk [11, 18], so this bound may not be sharp. Their proof technique handles Gaussian noise similarly, but requires the coordinates of the perturbed points to be independently distributed (and thus does not extend to uniform L^2 noise). The case where X consists of n copies of the same point corresponds to the classical question, in integral geometry, of counting the expected number of extreme points in a random sample of a domain $\mathcal{D}$. This number is $\Theta(k \log n)$ if $\mathcal{D}$ is a convex k-gon in the plane and $\Theta\left(n^{\frac{d-1}{d+1}}\right)$ if it is a ball in $\mathbb{R}^d$ [16, 17, 18] (see also [15] for a simple proof).

Contributions. We present near-tight bounds on the expected number of extreme points in a L^1, L^2 and L^∞ perturbation of n points in convex position that satisfy a certain "uniform density" condition. Let X be a set of n points on the unit sphere Γ (or any compact convex surface) in $\mathbb{R}^d$ such that there exist $\epsilon = \Theta\left(n^{\frac{1}{1-d}}\right)$ and $\kappa = O(1)$ such that any ball of radius ϵ centered on Γ contains between 1 and κ points of X. Let Y be a uniform L^p perturbation of X of amplitude δ, let $\mathrm{ex}(Y)$ denote the number of extreme points of Y and let $\mathbf{E}[V]$ denote the expected value of a random variable V. We prove the following bounds[1]:

- For $p = 2$, $\delta \in \left[\tilde{\Omega}\left(n^{\frac{2}{1-d}}\right), 1\right]$ and $d \geq 2$, $\mathbf{E}[\mathrm{ex}(Y)]$ is $\tilde{\Theta}\left((\sqrt{n})^{1-\frac{1}{d}}\left(\frac{1}{\sqrt[4]{\delta}}\right)^{d-\frac{1}{d}}\right)$ (Theorem 2).
- For $p = 1$ or ∞, $\delta \in \left[\tilde{\Theta}\left(\frac{1}{n^2}\right), \Theta(1)\right]$ and $d = 2$, $\mathbf{E}[\mathrm{ex}(Y)]$ is $\tilde{\Theta}\left(n^{\frac{1}{5}}\left(\frac{1}{\delta}\right)^{\frac{2}{5}}\right)$ (Theorem 7).

In particular, we refine the bound of Damerow and Sohler [5] for uniform L^∞ perturbation of convex point sets in the plane with "uniform density". Our bounds also show that already in the plane, the expected number of extreme points differs significantly between uniform L^2 and L^∞ perturbations. Note that the case $p = 1$ follows from the case $p = \infty$ by rotating the coordinate system by $\frac{\pi}{4}$.

Our proofs are based on a simple technique that we hope is applicable to a wider range of problems. We cover the space of directions $\mathbb{S}^{d-1}$ by m disks $D_1, \ldots, D_m$. For each disk D_i we define a pair of halfspaces that estimate the number of perturbed points that are extreme in some direction in that disk. In each pair, one halfspace, the *collector*, contains the other halfspace, the *witness*, and their respective position ensures that if the witness contains *at least one* perturbed point then the collector contains *all* perturbed points extreme in a direction of D_i. We prove that the expected number of extreme points of the perturbed set is $\tilde{\Theta}(m)$ if the following three conditions hold: (i) any region $\mathfrak{P}(x)$ intersects $O(1)$ distinct witnesses, (ii) the expected number of points in any witness is at least $\log n$, and (iii) the expected number of points in any

[1] We hide the logarithmic factors by writing $g(n) = \tilde{O}(f(n))$ if $g(n) = O(f(n)\log^\beta n)$ for some fixed β (and similarly for $\tilde{\Theta}()$ and $\tilde{\Omega}()$).

collector is $O(\log n)$. We then show that all three conditions can be met for the two types of perturbations we consider.

In dimension $d \leq 3$, the complexity of the convex hull is dominated by the number of extreme points, so Theorems 2 and 7 also bound the smoothed complexity of the convex hull for uniform L^2 and L^∞ perturbations. In higher dimension, the expected size of the convex hull corresponds to higher moments of the expected number of extreme points, and require a finer analysis.

We also compare the bounds given by Theorems 2 and 7 to two sets of experimental data. Our first dataset comes from simulations of random uniform L^2 and L^∞ perturbations. We observe that for n ranging from 10^3 up to 10^7, the constants hidden in the $\tilde{\Theta}$ seem very small (between 3 and 5) and the theoretical bounds are good estimates of the measured values (less than 5% error). Our second dataset comes from snap-rounding a regular n-gon on a regular grid of step δ. We observe that for n ranging from 10^3 up to 10^7 and $\delta \in \left[\frac{1000}{n^2}, \frac{\sqrt{n}}{100}\right]$, the number of extreme points is predicted (again, with less than 5% error) by $4\delta^{-\frac{2}{5}} n^{\frac{1}{5}}$, that is our bound for uniform L^∞ noise. This provides empirical evidence that the smoothed analysis model can help predict the complexity for data represented with finite precision.

2 Outline of the method

In this section, we introduce our perturbation models and describe our approach to proving bounds on the expected number of extreme points after perturbation.

Perturbations. A *perturbation method* in $\mathbb{R}^d$ is a map $\mathfrak{P} : \mathbb{R}^d \to D(\mathbb{R}^d)$, where $D(\mathbb{R}^d)$ denotes the set of probability distributions over $\mathbb{R}^d$. Recall that the *support* of a probability distribution μ is the smallest closed subset $U \subset \mathbb{R}^d$ such that $\mu(U) = 1$. A *perturbation of* a point set $X = \{x_1, \ldots, x_n\}$ by (the perturbation method) $\mathfrak{P}$ is a set of n random points $\mathfrak{P}(X) = \{y_1, \ldots, y_n\}$ where y_i is chosen according to the probability distribution $\mathfrak{P}(x_i)$.

In this paper, we consider perturbation methods where $\mathfrak{P}(x)$ is a uniform probability on a compact (convex) region around x whose size is governed by some parameter δ. Specifically, in Section 3 we let $\mathfrak{P}(x)$ be the uniform distribution in a ball of radius δ centered at x, and in Section 4 we let $\mathfrak{P}(x)$ be the uniform distribution in an axis-parallel square of side δ centered at x. When clear from the context, we identify $\mathfrak{P}(x)$ with its support and say that $\mathfrak{P}(x)$ is either a ball or a square. In each case, we are interested in the following question: given a set X of n points in convex position, how does the expected number of extreme points of $\mathfrak{P}(X)$ depend on n and δ?

The general bound. We say that a pair (C, W) of halfspaces in $\mathbb{R}^d$ are *nested* if the first contains the second: $W \subset C$; we call C the *collector* and W the *witness*. A nested pair (C, W) *controls* a direction $\overrightarrow{u} \in \mathbb{S}^{d-1}$ for a set X and a perturbation method $\mathfrak{P}$ if whenever $W \cap \mathfrak{P}(X) \neq \emptyset$ the point extremum in direction $\overrightarrow{u}$ belongs to C. In other words, if the witness W contains a perturbed point, the collector C collects all points extreme in all directions controlled by (C, W). We say that a family of nested pairs controls a direction if some pair in

the family does. Given a region $R \subset \mathbb{R}^d$, we denote by $|R \cap \mathfrak{P}(X)|$ the expected number of points of a perturbation $\mathfrak{P}(X)$ of X that are contained in R.

Lemma 1. *Let X be a set of n points in $\mathbb{R}^d$, $\mathfrak{P}$ a perturbation method and $N = \{(C_i, W_i)\}_{1 \leq i \leq m}$ a family of nested pairs of halfspaces that (a) controls all directions in $\mathbb{S}^{d-1}$ and (b) has only $O(1)$ witnesses that meet any given support $\mathfrak{P}(x)$ with $x \in X$. Then, the expected number $\mathbf{E}[\mathrm{ex}(\mathfrak{P}(X))]$ of extreme points in a perturbation of X is:*

$$\begin{aligned} \Omega\left(m - \sum_{1 \leq i \leq m} e^{-|W_i \cap \mathfrak{P}(X)|}\right) &\leq \mathbf{E}[\mathrm{ex}(\mathfrak{P}(X))] \\ &\leq n \sum_{1 \leq k \leq m} e^{-|W_i \cap \mathfrak{P}(X)|} + \sum_{1 \leq k \leq m} |C_i \cap \mathfrak{P}(X)|. \end{aligned}$$

Proof. To save breath, we denote the perturbed point set by $Y = \mathfrak{P}(X)$. For a witness or a collector H, we let $\alpha(H) = |H \cap Y|$ and denote by $\beta(H)$ the probability that $H \cap Y$ is empty. If we write t_k the probability that the k^{th} point of Y is in H, we have

$$\alpha(H) = \sum_{1 \leq k \leq n} t_k \quad \text{and} \quad \beta(H) = \prod_{1 \leq k \leq n} (1 - t_k),$$

and the classical inequality $1 - t \leq e^{-t}$ yields $\beta(H) \leq e^{-\alpha(H)}$.

If a witness halfspace W_i is non-empty, it contains the point of $\mathfrak{P}(X)$ that is extremal in the direction of its inner tangent. Condition (b) ensures that any extremal point is contained in $O(1)$ witness halfspaces. Thus, the number of extreme points of Y is at least a constant times the number of non-empty witness halfspaces, and the lower bound follows. Now, let $\mathrm{ex}_k(Y)$ denote the number of points in Y that are extreme in a direction controlled by (C_k, W_k). If $Y \cap W_k$ is non-empty then all points extreme in such directions are in C_k, and $\mathrm{ex}_k(Y) \leq \alpha(C_k)$; otherwise, $\mathrm{ex}_k(Y)$ is at most n. Altogether, we get

$$\mathbf{E}[\mathrm{ex}_k(Y)] \leq n\beta(W_k) + (1 - \beta(W_k))\alpha(C_k) \leq e^{-\alpha(W_k)}n + \alpha(C_k),$$

and Condition (a) implies that $\mathrm{ex}(Y) \leq \sum_{1 \leq k \leq m} \mathrm{ex}_k(Y)$. The upper bound follows. □

Applications. Applying Lemma 1 essentially boils down to making sure that most witnesses are non-empty (which requires large $|W_i \cap \mathfrak{P}(X)|$) while most collectors collect few points (which requires small $|C_i \cap \mathfrak{P}(X)|$). We will find trade-offs where $|W_i \cap \mathfrak{P}(X)|$ is greater than $\log n$ while $|C_i \cap \mathfrak{P}(X)|$ remains $\Theta(\log n)$. Plugged into Lemma 1, these bounds will ensure that $\mathbf{E}[\mathrm{ex}(\mathfrak{P}(X))] = \tilde{\Theta}(m)$. For the sake of simplicity, we unfold this method in the case where X is a "nice" sample of the unit sphere Γ. Given a halfspace H (witness or collector), we will need to estimate $|H \cap \mathfrak{P}(X)|$. Letting h_x represents the distance between ∂H and the point in $H \cap \mathfrak{P}(x)$ furthest away from ∂H and writing $\mathsf{area}(h_x)$ for the volume of $H \cap \mathfrak{P}(x)$, we have $|H \cap \mathfrak{P}(X)| = \sum_{x \in X} \mathsf{area}(h_x)$. We will then express $\mathsf{area}(h_x)$ with respect to the depth h of H defined by $h = \max_{x \in \Gamma} h_x$. It then remains to estimate the depths c_i and h_i of the nested halfplanes C_i and W_i needed to satisfy conditions (a) and (b) in Lemma 1.

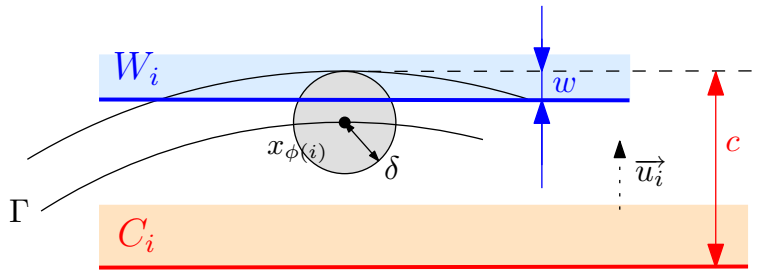


Figure 1: Notations for witnesses and collectors.

3 Perturbations when the support is a ball

Recall that an (ϵ, κ)-*sample* of a set Γ is a point set $X \subset \Gamma$ such that any ball of radius ϵ centered on Γ contains between 1 and κ points from the sample X. In this section, we prove the following

Theorem 2. *Let $\delta \in \left[\tilde{\Omega}\left(n^{\frac{2}{1-d}}\right), 1\right]$. Let X be a $\left(\Theta\left(n^{\frac{1}{1-d}}\right), \Theta(1)\right)$-sample of size n of the unit sphere in $\mathbb{R}^d$ and let $\mathfrak{P}(x)$ be the uniform distribution on the ball with center x and radius δ. The expected number of extreme points of $\mathfrak{P}(X)$ is $\tilde{\Theta}\left((\sqrt{n})^{1-\frac{1}{d}}\left(\frac{1}{\sqrt[4]{\delta}}\right)^{d-\frac{1}{d}}\right)$.*

We first prove Theorem 2 in the plane, and then extend the technique to arbitrary dimension.

3.1 The two-dimensional case

Let $\delta > 0$ and let $\mathfrak{P}(x)$ be the uniform distribution on the disk with center x and radius δ. We denote by Γ the unit circle centered at the origin O in $\mathbb{R}^2$ and let $X = \{x_1, \ldots, x_n\}$ be a $\left(\Theta\left(\frac{1}{n}\right), \Theta(1)\right)$-sample of Γ.

Witnesses and collectors. We first define our family of nested pairs using three parameters: m counts the number of pairs, c controls the depth of the collectors and w that of the witnesses. We then give conditions on m, c and w that guarantee that the family $\{(C_i, W_i)\}_{1 \leq i \leq m}$ meets the conditions of Lemma 1.

We pick m points $x_{\phi(1)}, \ldots, x_{\phi(m)}$ in X with indices as equally spaced as possible and let $\overrightarrow{u_i} = \overrightarrow{Ox_{\phi(i)}}$. We then define W_i (resp. C_i) as the halfplane with inner normal $\overrightarrow{u_i}$ and whose boundary passes through the point $O + (1 + \delta - w)\overrightarrow{u_i}$ (resp. $O + (1 + \delta - c)\overrightarrow{u_i}$). We denote the resulting family of nested halfplanes by $N(m, w, c)$ (see Figure 1).

Lemma 3. *For any $\delta \in [0, 1]$, there exists a positive real number $w_0 = \Theta\left(\frac{1}{m^2}\right)$ such that for any w satisfying $w_0 \leq w = O\left(\frac{1}{m^2}\right)$ and for any $c \geq 9w$ the family $N(m, w, c)$ controls all directions in $\mathbb{S}^1$ and has only $O(1)$ witnesses meeting any given support $\mathfrak{P}(x)$ with $x \in X$.*

Proof. Let (C, W) be a nested pair in $N(m, w, c)$. Let Λ denote the circle of radius $1+\delta$ centered in O. We define $\eta(W)$ as the directions corresponding to the arc $\Lambda \cap W$:

$$\eta(W) = \left\{ \frac{\overrightarrow{OA}}{\|\overrightarrow{OA}\|} \mid A \in \Lambda \cap W \right\},$$

Let w_1 and w_2 denote the two intersection points of ∂W with Λ. The line through w_1 with normal Ow_2 meets the circle Λ in another point $t_1 \neq w_1$. Define t_2 similarly.

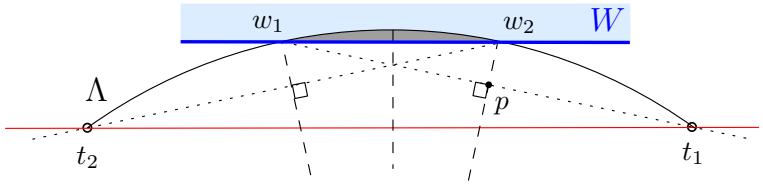


Figure 2: The points w_i and t_i.

Since the chord w_1t_1 is perpendicular to the radial line Ow_2, their intersection point is the middle of the segment w_1t_1. It follows that the angles $\widehat{w_1Ow_2}$ and $\widehat{w_2Ot_1}$ are equal. Similarly, we have $\widehat{w_1Ow_2} = \widehat{t_2Ow_1}$. For $w \leq (1+\delta)(1-\frac{\sqrt{3}}{2})$, we have $\widehat{w_1Ow_2} \leq \frac{\pi}{3}$ and the line t_1t_2 intersects the segment connecting O to the midpoint of w_1w_2. We let $O + (1+\delta-\tau)u_k$ be this intersection point. Any point in W lies in the shaded region depicted in Figure 2, and thus dominates any point inside Λ and below (t_1t_2), for all directions in $\eta(W)$. Since $\mathfrak{P}(X)$ lies inside Λ, we get that if $c \geq \tau$ the pair (C, W) controls any direction in $\eta(W)$. Setting $\theta = \frac{\widehat{w_1Ow_2}}{2}$, we have by construction $3\theta = \frac{\widehat{t_1Ot_2}}{2}$. It follows that

$$1+\delta-\tau = (1+\delta)\cos 3\theta \quad \text{and} \quad 1+\delta-w = (1+\delta)\cos\theta,$$

and, consequently, $\frac{\tau}{e} = \frac{1-\cos 3\theta}{1-\cos\theta} \leq 9$. Hence, if $c \geq 9w$ the pair (C, W) controls any direction in $\eta(W)$.

The sampling condition on X ensures that every angle $\widehat{x_{\phi(l)}Ox_{\phi(l+1)}}$ is $\Theta(\frac{1}{m})$. On the other hand, the opening angle of the arc $\eta(W)$ is $2\arccos\left(1-\frac{w}{1+\delta}\right)$, which is $\Theta(\sqrt{w})$ since $\delta \leq 1$. By choosing w larger than a positive real number $w_0 = \Omega(\frac{1}{m^2})$, we force the sets $\eta(W_1), \ldots, \eta(W_m)$ to cover $\mathbb{S}^1$. Also, the number of sectors $x_{\phi(l)}x_{\phi(l+1)}$ that can meet a given arc $\Lambda \cap W$ is $O(m\sqrt{w})$. It follows that for any choice of w that is $O(\frac{1}{m^2})$, the number of witnesses intersected by any given $\mathfrak{P}(x)$ is $O(1)$. □

Expected number of points in a halfspace. We now estimate the expected number of points in $H \cap \mathfrak{P}(X)$ where H designates either a witness or a collector halfplane. Since both types of halfplanes are defined similarly, this can be done in a single stroke as follows.

Lemma 4. *Let $x \in X$, $0 \leq h \leq \delta$ and let H be the halfplane with normal $\overrightarrow{u} = \overrightarrow{Ox}$ passing through the point $O + (1+\delta-h)\overrightarrow{u}$. The expected size of $H \cap \mathfrak{P}(X)$ is $\Theta\left(nh^2\delta^{-\frac{3}{2}}\right)$.*

Proof. Let z_j denote a vertex of X that is j steps away from $x = z_0$. If we denote the distance from ∂H to z_j by $\delta - h_j$ we have $h_j = h - (1 - \cos \widehat{xOz_j})$ (see Figure 3-left).

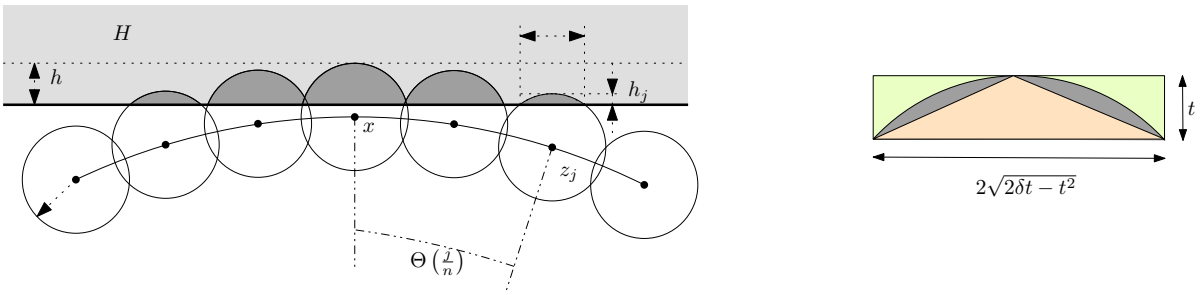


Figure 3: (Left): the distance between ℓ and x_j. (Right): estimation of $\mathsf{area}(t)$.

The sampling condition on X yields that $\widehat{xOz_j} = \Theta\left(\frac{j}{n}\right)$ and since $\forall \theta \in [0, \frac{\pi}{2}]$ $\frac{\theta^2}{3} \leq 1 - \cos\theta \leq \frac{\theta^2}{2}$ we have $h_j = h - \Theta\left(\frac{j^2}{n^2}\right)$. The probability that $\mathfrak{P}(z_j) \in H$ then writes as $\frac{\mathsf{area}(h_j)}{\pi\delta^2}$, where $\mathsf{area}(t)$ denotes the smallest portion of a disk of radius δ cut by a line ℓ that passes at distance $\delta - t$ from its center. To estimate $\mathsf{area}(t)$, we refer to Figure 3-right. Let s denote the length of the chord intercepted by ℓ. From $s = 2\sqrt{\delta^2 - (\delta - t)^2} = 2\sqrt{2\delta t - t^2}$ and $0 < t < \delta$, we get that $2\sqrt{\delta t} < s < 2\sqrt{2\delta t}$. Since the small portion of the disk contains a triangle with base s and height t and is contained in a rectangle with the same side length. We have that $\sqrt{\delta} t^{3/2} < \mathsf{area}(t) < \sqrt{\delta}(2t)^{3/2}$, and $\mathsf{area}(t) = \Theta(\sqrt{\delta} t^{3/2})$. Summing over all points z_j such that $\mathfrak{P}(z_j)$ intersects H, we get[2]

$$
\begin{aligned}
|H \cap \mathfrak{P}(X)| &= \sum_{h_j \geq 0} \frac{\mathsf{area}(h_j)}{\pi\delta^2} \\
&= \frac{\mathsf{area}(h)}{\pi\delta^2} + 2 \sum_{j=1}^{\Theta(n\sqrt{h})} \frac{\mathsf{area}(h_j)}{\pi\delta^2} = \Theta\left(\frac{\sqrt{\delta} h^{3/2}}{\pi\delta^2}\right) + \sum_{j=1}^{\Theta(n\sqrt{h})} \Theta\left(\frac{\sqrt{\delta} h_j^{3/2}}{\pi\delta^2}\right) \\
&= \delta^{-\frac{3}{2}} \Theta\left(h^{\frac{3}{2}} + \sum_{j=1}^{\Theta(n\sqrt{h})} h_j^{\frac{3}{2}}\right) \\
&= \delta^{-\frac{3}{2}} \Theta\left(h^{\frac{3}{2}} + \sum_{j=1}^{\Theta(n\sqrt{h})} \left(h - \Theta\left(\frac{j^2}{n^2}\right)\right)^{3/2}\right) \\
&= \Theta\left(nh^2\delta^{-\frac{3}{2}}\right),
\end{aligned}
$$

which proves the statement. □

Wrapping up. We can now prove Theorem 2 for $d = 2$. We choose $m = \Theta\left(\delta^{-3/8} n^{1/4} \log^{-1/4} n\right)$, $w = \Theta\left(\delta^{3/4} n^{-1/2} \log^{1/2} n\right)$ and $c = 9w$, and con-

[2] For the last equality, we use the fact that for $\forall a > 1, \forall b \geq 0$, we have $\sum_{j=1}^{p} (1 - (j/p)^a)^b = \Theta(p)$.

sider the family $N(m,c,w) = \{(C_i, W_i)\}_{1\le i\le m}$ of nested pairs defined above. Lemma 4 yields that

$$|W_i \cap \mathfrak{P}(X)| = \Theta\Big(nw^2\delta^{-\frac{3}{2}}\Big) = \Theta\bigg(n\Big(\delta^{3/4}n^{-1/2}\log^{1/2} n\Big)^2\delta^{-\frac{3}{2}}\bigg) = \Theta(\log n)\,,$$

and, similarly, we have $|C_i \cap \mathfrak{P}(X)| = \Theta(\log n)$. We choose the constant in the $\Theta(\cdot)$ that defines w so that $\forall i, |W_i \cap \mathfrak{P}(X)| \ge \log n$; $|C_i \cap \mathfrak{P}(X)|$ then remains $\Theta(\log n)$. Now, Lemma 3 ensures that the general bound of Lemma 1 applies for $N(m,c,w)$, and we have:

$$\Omega\left(m - \sum_{1\le i\le m} e^{-|W_i\cap\mathfrak{P}(X)|}\right) \;\le\; \mathbf{E}[\mathrm{ex}(\mathfrak{P}(X))] \;\le\; n\sum_{1\le k\le m} e^{-|W_i\cap\mathfrak{P}(X)|} + \sum_{1\le k\le m}|C_i\cap\mathfrak{P}(X)|.$$

This becomes

$$\Omega\bigg(m\Big(1-\frac{1}{n}\Big)\bigg) \le \mathbf{E}[\mathrm{ex}(\mathfrak{P}(X))] \le O(m(1+\log n))\,,$$

which rewrites as $\mathbf{E}[\mathrm{ex}(\mathfrak{P}(X))] = \tilde{\Theta}(m)$. The statement follows.

3.2 Arbitrary dimension

Let us now generalize the previous arguments to arbitrary dimension. Let Γ be the unit sphere in $\mathbb{R}^d$ and X a $\Big(\Theta\Big(n^{\frac{1}{1-d}}\Big), \Theta(1)\Big)$-sample of Γ. Let $m \le n$ be an integer and let w and c be two positive reals. We pick a subset X_ϕ of X of size m that forms a $\Big(\Theta\Big(m^{\frac{1}{1-d}}\Big), \Theta(1)\Big)$-sample of Γ and associate a nested pair (C_x, W_x) to each point $x \in X_\phi$. Specifically, W_x is the halfspace with inner normal $\overrightarrow{Ox}$ whose boundary hyperplane passes through the point $O + (1+\delta-w)\overrightarrow{Ox}$; C_x is defined similarly, with the parameter w replaced by c. Lemma 3 easily generalizes as follows:

Lemma 5. *For any $\delta \in [0,1]$, there exists a positive real number $w_0 = \Theta\Big(m^{\frac{2}{1-d}}\Big)$ such that for any w satisfying $w_0 \le w = O\Big(m^{\frac{2}{1-d}}\Big)$ and for any $c \ge 9w$ the family $\{(C_x, W_x)\}_{x\in X_\phi}$ controls all directions in $\mathbb{S}^{d-1}$ and has only $O(1)$ witnesses meeting any given support $\mathfrak{P}(x)$ with $x \in X$.*

Proof. Let Λ denote the sphere of radius $1+\delta$ centered in O. Let $\eta(H)$ denote the set of directions $\overrightarrow{OA}$ where $A \in H \cap \Lambda$. Let $x \in X_\phi$. The fact that if $c \ge 9w$ then a non-empty $W_x \cap \mathfrak{P}(X)$ guarantees that any point of $\mathfrak{P}(X)$ extreme in a direction $\overrightarrow{u} \in \eta(W_x)$ must belong to C_x is clear if $\overrightarrow{u} // \overrightarrow{Ox}$ and follows, in the other cases, from considering the projection on the plane spanned by O, x and $\overrightarrow{u}$ and applying the arguments from the two-dimensional case. The portion of Λ cut out by W_x has diameter $\Theta(\sqrt{w})$ as in the two-dimensional case. Since its $(d-1)$-volume is $\Theta\big(\frac{1}{m}\big)$, the caps $\{\eta(W_x)\}_{x\in X_\phi}$ cover $\mathbb{S}^{d-1}$ for $\sqrt{w}^{d-1} = \Theta\big(\frac{1}{m}\big)$.

It remains to discuss Condition (b) of Lemma 1. Given $x \in X_\phi$, and $y \in X$, let $x' = Ox \cap \Lambda$ and $y' = Oy \cap \Lambda$. y meet W_x if there is a point $z \in W_x \cap \mathfrak{P}(x)$. Since the portion of Λ cut out by W_x has diameter $\Theta(\sqrt{w})$, we have

$|x'z| \leq \Theta(\sqrt{w})$, similarly the portion of Λ cut out by the plane orthogonal to Oy through z has diameter $O(\sqrt{w})$, and thus $|y'z| \leq \Theta(\sqrt{w})$. Finally we get $xy \leq x'y' \leq \Theta(\sqrt{w})$, then the fact that X_ϕ is a $\left(\Theta\left(m^{\frac{1}{1-d}}\right), \Theta(1)\right)$-sample ensures condition (b). □

Estimation. Lemma 4 generalizes as follows.

Lemma 6. *Let $x \in X$, $0 \leq h \leq \delta$ and let H be the halfplane with normal $\overrightarrow{u} = \overrightarrow{Ox}$ passing through the point $O + (1 + \delta - h)\overrightarrow{u}$. The expected size of $H \cap \mathfrak{P}(X)$ is $\Theta\left(nh^d\delta^{-\frac{d+1}{2}}\right)$.*

Proof. Let $z \in X$ be a point such that $\mathfrak{P}(z)$ intersects H. We let $\theta_z = \widehat{xOz}$ and denote the distance from ∂H to z by $\delta - h_z$. Observe that the relation $h_z = h - \Theta(\theta_z^2)$ established in the proof of Lemma 4 remains valid. We then proceed as in the proof of Lemma 4, summing up the volumes of $H \cap \mathfrak{P}(z)$, with the following two changes. First, we have to replace the area function $\mathsf{area}(t)$ by a volume function $\mathsf{volume}(t)$. We can, as in 2D, sandwich $H \cap \mathfrak{P}(z)$ between a circular pyramid (inside) and a box (outside), both of height t and basis of volume $\Theta\left((\sqrt{\delta t})^{d-1}\right)$. We thus have $\mathsf{volume}(t) = \Theta\left(\delta^{\frac{d-1}{2}} t^{\frac{d+1}{2}}\right)$. Second, we now have to sum the contributions of all the points of X lying in a $(d-1)$-dimensional circular cap. Let Γ_H denote the cap of Γ consisting of points p for which $\mathfrak{P}(p)$ meets H. Observe that Γ_H is centered in x and let $\theta(m)$ denote its opening angle. We subdivide the interval $[0, \theta(m)]$ into $\Theta\left(n^{\frac{1}{d-1}}\sqrt{h}\right)$ sub-intervals of width $n^{\frac{1}{1-d}}$. Since X is a $\left(\Theta\left(n^{\frac{1}{1-d}}\right), \Theta(1)\right)$-sample of Γ, the number of points $z \in X$ with $\theta_z \in [\theta, \theta + n^{\frac{1}{1-d}}]$ is $\Theta\left(\theta^{d-2} n^{\frac{d-2}{d-1}}\right)$. The sum then becomes

$$
\begin{aligned}
|H \cap \mathfrak{P}(X)| &= \sum_{h_z \geq 0} \frac{\mathsf{volume}(h_z)}{\Theta(\delta^d)} \\
&= \Theta(1) \sum_{i=0}^{\Theta(n^{\frac{1}{d-1}}\sqrt{h})} \sum_{z;\theta_z \in [i \cdot n^{1-d}, (i+1)n^{1-d}]} \frac{\mathsf{volume}(h_z)}{\delta^d} \\
&= \Theta(1) \sum_{i=0}^{n^{\frac{1}{d-1}}\sqrt{h}} \frac{\left(i \cdot n^{\frac{1}{1-d}}\right)^{d-2}}{\left(n^{\frac{1}{1-d}}\right)^{d-2}} \frac{\delta^{\frac{d-1}{2}}\left(h - i^2 n^{\frac{2}{1-d}}\right)^{\frac{d+1}{2}}}{\delta^d} \\
&= \Theta(1)\left(\frac{\delta^{\frac{d-1}{2}}}{\delta^d} h^{\frac{d+1}{2}}\right) \sum_{i=0}^{n^{\frac{1}{d-1}}\sqrt{h}} i^{d-2} \\
&= \Theta(1)\delta^{-\frac{d+1}{2}} h^{\frac{d+1}{2}} \left(n^{\frac{1}{d-1}}\sqrt{h}\right)^{d-1} \\
|H \cap \mathfrak{P}(X)| &= \Theta\left(n \cdot h^d \cdot \delta^{-\frac{d+1}{2}}\right),
\end{aligned}
$$

which proves the statement. □

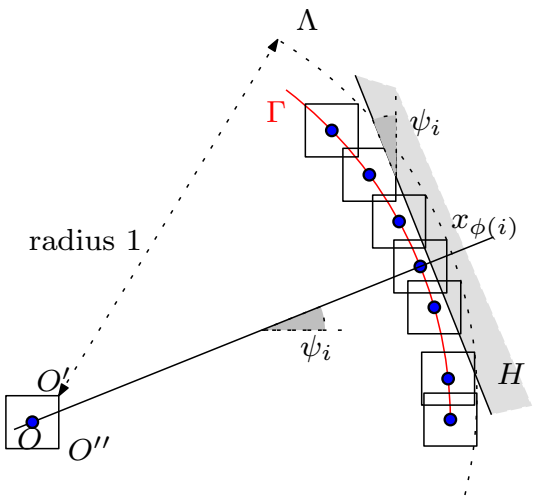


Figure 4: Support of the perturbation is an axis parallel square

Wrapping-up. We can now prove Theorem 2 in all dimensions. We choose $m = \Theta\Big(\delta^{\frac{1-d^2}{4d}} n^{\frac{d-1}{2d}} \log^{\frac{1-d}{2d}} n\Big)$ and let $w = \Theta\Big(m^{\frac{2}{1-d}}\Big) = \Theta\Big(\delta^{\frac{d+1}{2d}} n^{-\frac{1}{d}} \log^{\frac{1}{d}} n\Big)$ and $c = 9w$. Lemma 6 then implies that

$$|\mathfrak{P}(x) \cap W_i| \geq \log n \quad \text{and} \quad |\mathfrak{P}(x) \cap C_i| = \Theta(\log n)\,,$$

for a suitable choice of constant in the $\Theta(\cdot)$ that defines m. For n large enough, Lemma 5 guarantees that we can use the family $\{(C_i, W_i)\}_{1 \leq i \leq m}$ in Lemma 1, and we get:

$$\Omega\left(m\left(1 - \frac{1}{n}\right)\right) \leq \mathbf{E}[\mathrm{ex}(P)] \leq O(m \log n)\,.$$

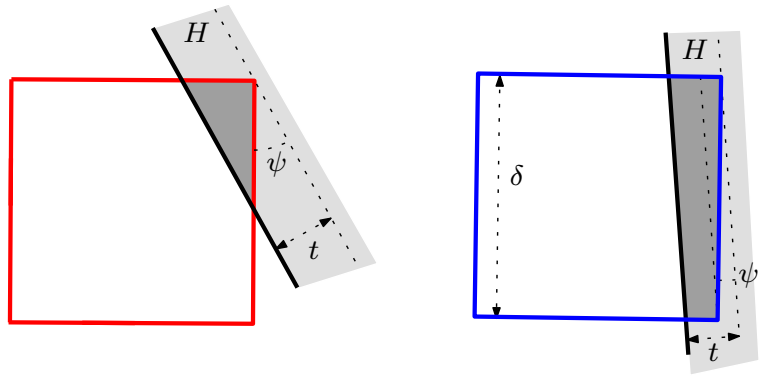
This rewrites as $\mathbf{E}[\mathrm{ex}(P)] = \tilde{\Theta}(m)$ and the statement follows.

4 Perturbations when the support is a square

The construction in Section 3 can be adapted to perturbation methods whose supports are hypercubes. However, this leads to more technical considerations so we will restrict ourselves to the two-dimensional case (see Figure 5). We state our main result:

Theorem 7. *Let $\delta \in \left[\tilde{\Theta}\left(\frac{1}{n^2}\right), \Theta(1)\right]$. Let X be a $\left(\Theta\left(\frac{1}{n}\right), \Theta(1)\right)$-sample of size n of the unit circle in $\mathbb{R}^2$ and let $\mathfrak{P}(x)$ be the uniform distribution in the axis-parallel square with center x and side length δ. The expected number of extreme points of $\mathfrak{P}(X)$ is $\tilde{\Theta}\Big(n^{\frac{1}{5}}\left(\frac{1}{\delta}\right)^{\frac{2}{5}}\Big)$.*

Let $\mathfrak{P}(x)$ denote the square of side δ centered in x and parallel to the axes. Due to symmetry, in most of the proof below we will only consider the first octant of the circle Γ.



one corner behavior
$t < \delta \sin \psi$
$\mathsf{area}(t) = \Theta(t^2 \psi^{-1})$

two corners behavior
$t > \delta \sin \psi$
$\mathsf{area}(t) = \Theta(\delta t)$

Figure 5: One (left) and two (right) corners behavior

4.1 Collectors and witnesses.

We define our family of nested pairs by picking m points $x_{\phi(1)}, \ldots, x_{\phi(m)}$ in X in a way that will be specified later. Specifically, we associate to each point $x_{\phi(i)}$ a nested pair of half-planes C_i and W_i whose common inner normal is $\overrightarrow{u_i} = \overrightarrow{Ox_{\phi(i)}}$ and parameterize the location of C_i and W_i as follows. Due to symmetry, we only explain the parameterization when $x_{\phi(i)}$ belongs to the first quadrant. Observe that as a point x moves on the boundary of Γ, the upper right corner of the square $\mathfrak{P}(x)$ moves on the circle Λ with radius 1 and center O' (see figure left above). In particular, the upper right corner of $\mathfrak{P}(x)$ traces a piece of the boundary of $\bigcup_{x \in \Gamma} \mathfrak{P}(x)$. We let w_i and c_i be two real numbers such that the boundary of W_i (resp. C_i) passes through $O' + (1 - w_i)\overrightarrow{u_i}$ (resp. $O' + (1 - c_i)\overrightarrow{u_i}$). Hence, the family of nested pairs is described by m, the map ϕ and the $2m$ parameters $\{w_i\}_{1 \leq i \leq m}$ and $\{c_i\}_{1 \leq i \leq m}$. Afterwards, we will set $c_i = 9w_i$ and adjust the depth w_i, depending on the angle ψ_i between $\overrightarrow{u_i}$ and the abscissa axis.

4.2 Two corners and one corner behaviors

Given a square $\mathfrak{P}(x)$ and a halfplane H, we denote by $\psi \in [0, \frac{\pi}{4}]$ the smallest angle between the boundary of H and one side of $\mathfrak{P}(x)$. We express the area of $H \cap \mathfrak{P}(x)$ in terms of the distance t between ∂H and the corner of $H \cap \mathfrak{P}(x)$ furthest away from ∂H. We distinguish two cases, depending on whether H contains one or two corners of $\mathfrak{P}(x)$ (see Figure 5):

the *one corner behavior:*
$$t \leq \delta \sin \psi\text{:} \quad \mathsf{area}(t) = t^2(\tan \psi + \cot \psi) = \Theta(t^2 \psi^{-1}).$$
the *two corners behavior:*
$$t \geq \delta \sin \psi\text{:} \quad \mathsf{area}(t) = \frac{t\delta}{\cos \psi} = \Theta(t\delta).$$

4.3 Nested pairs with the one corner behavior

Lemma 8. *There exists* $\psi_\square = \Theta(n^{-\frac{2}{3}}\delta^{-\frac{1}{3}}\log^{\frac{2}{3}} n)$ *such that if* $\psi_i \geq \psi_\square$ *and* $w_i \geq \delta^{\frac{4}{5}}\psi_i^{\frac{2}{5}} n^{-\frac{2}{5}}\log^{\frac{2}{5}} n$, *then* $|W_i \cap \mathfrak{P}(X)| = \Theta\Big(nw_i^{\frac{5}{2}}\delta^{-2}\psi_i^{-1}\Big)$ *and* $|C_i \cap \mathfrak{P}(X)| = \Theta\Big(nc_i^{\frac{5}{2}}\delta^{-2}\psi_i^{-1}\Big)$.

Proof. Let $H \in \{W_i, C_i\}$ and define $h \in \{w_i, c_i\}$ such that the boundary of H passes through $O' + (1-h)\overrightarrow{u_i}$. Suppose H contains at most one corner of any square $\mathfrak{P}(x)$ for $x \in X$. Then, noticing that the smallest angle between the boundary of H and one side of any $\mathfrak{P}(x)$ is ψ_i and mimicking the computation of $H \cap \mathfrak{P}(X)$ in Lemma 4, we get:

$$|H \cap \mathfrak{P}(X)| = \Theta\left(\sum_{j=0}^{n\sqrt{h}} \left(h - \frac{j^2}{n^2}\right)^2 \frac{\psi_i^{-1}}{\delta^2}\right) = \Theta\Big(nh^{\frac{5}{2}}\delta^{-2}\psi_i^{-1}\Big).$$

It remains to check that all squares intersecting H are indeed in the one corner behavior. For this, it is enough to check that the square centered at $x_{\phi(i)}$ has only one corner in C_i. A sufficient condition is that $c_i \leq \delta \sin\psi_i$ which holds if $\psi_i \geq \psi_\square$ for some $\psi_\square = \Theta(\delta^{-\frac{1}{3}} n^{-\frac{2}{3}}\log^{\frac{2}{3}} n)$. □

4.4 Nested pairs with the two corners behavior

Lemma 9. *There exists* $w_\square = \Theta(n^{-\frac{2}{3}}\delta^{\frac{2}{3}}\log^{\frac{2}{3}} n)$ *such that if* $\psi_i \leq \psi_\square$ *and* $w_i = w_\square$, *then* $|W_i \cap \mathfrak{P}(X)| > \log n$ *and* $|C_i \cap \mathfrak{P}(X)| = \Theta(\log n)$.

Proof. Let $h = \Theta(n^{-\frac{2}{3}}\delta^{\frac{2}{3}}\log^{\frac{2}{3}} n)$ and consider the half-plane H whose boundary passes through $O'+(1-h)\overrightarrow{u_i}$. To bound $|H \cap \mathfrak{P}(X)|$, we need to take into account that some of the squares $\mathfrak{P}(x)$ for $x \in X$ intersect H with the two corners behavior. Let z_j be the point in X that is j steps away from $x_{\phi(i)} = z_0$. Let h_j denote the distance between ∂H and the corner of $\mathfrak{P}(z_j)$ in H furthest away from ∂H. Similar to the disk case, we have $h_j = h - \Theta\Big(\frac{j^2}{n^2}\Big)$. As we consider $\mathfrak{P}(z_j)$ for increasing values of $|j|$, the upper right corner (which corresponds to the corner in H furthest away from ∂H) moves on circle Λ and exits H when $|j|$ reaches some value $j_\diamond$. The lower right corner moves on a circle with radius 1 and center O'' (see the figure at the beginning of this appendix) and exits H for $|j| = j_\square \leq j_\diamond$.

- To evaluate $j_\diamond$, we set $h_{j_\diamond} = 0$ that is $h - \Theta\Big(\frac{j_\diamond^2}{n^2}\Big) = 0$ which gives $j_\diamond = \Theta(n\sqrt{h})$ as in the disk case.
- To evaluate $j_\square$, we set $h_{j_\square} = \delta\sin\psi_i$. Since $\delta\sin\psi_i = O(n^{-\frac{2}{3}}\delta^{\frac{2}{3}}\log^{\frac{2}{3}} n) = O(h)$, we get that $h - \Theta\Big(\frac{j_\square^2}{n^2}\Big) = \delta\sin\psi_i$ gives $j_\square = \Theta(n\sqrt{h})$.

Hence, $j_\square \leq j_\diamond$ are of the same order. Now, observe that if a square $\mathfrak{P}(z_j)$ intersects H with the one corner behavior, the area of the intersection $H \cap \mathfrak{P}(z_j)$ is upper bounded by the formula given for the two corner behavior, *i.e.* $\mathsf{area}(h_j) \leq \delta h_j/\cos\psi_i$. Using this observation, we deduce upper and lower

bounds on $|H \cap \mathfrak{P}(X)|$:

$$\sum_{0\le j\le j_\square} \frac{\Theta(h_j\delta)}{\delta^2} \;\le\; |H\cap\mathfrak{P}(X)| = \sum_{h_j\ge 0} \frac{\mathsf{area}(h_j)}{\delta^2} \;\le\; \sum_{0\le j\le j_\diamond} \frac{\Theta(h_j\delta)}{\delta^2}.$$

It follows that:

$$|H\cap\mathfrak{P}(X)| \;=\; \Theta\left(\sum_{j=0}^{\Theta(n\sqrt{h})} \frac{h_j}{\delta}\right) \;=\; \Theta\left(\frac{1}{\delta}\sum_{j=0}^{n\sqrt{h}}\left(h-\frac{j^2}{n^2}\right)\right) \;=\; \Theta\Big(nh^{\frac{3}{2}}\delta^{-1}\Big).$$

Note that we can always choose the constant in the $\Theta(\cdot)$ that defines h so that $|H\cap\mathfrak{P}(X)| > \log n$. Denoting the corresponding value of h by $w_\square$, we get the claimed bounds on $|W_i\cap\mathfrak{P}(X)|$ and $|C_i\cap\mathfrak{P}(X)|$. □

4.5 Choosing the nested pairs

Within the first octant, we partition points $x_{\phi(i)}$ into three groups and for each group, associate a family of nested pairs. Specifically:

-1- Let $N_\square = \{(W_i, C_i)\}_{1\le i\le m_\square}$ be a family of nested pairs defined as follows. Recall $\psi_\square = \Theta(n^{-\frac{2}{3}}\delta^{-\frac{1}{3}}\log^{\frac{2}{3}} n)$ and place nested pairs with the parameters of Lemma 9, that is $w_i = w_\square = \Theta(n^{-\frac{2}{3}}\delta^{\frac{2}{3}}\log^{\frac{2}{3}} n)$. We take $\phi(i) = i\cdot n\cdot\sqrt{w_\square}$. Using the sampling condition, we have $\psi_i = \Theta\Big(\frac{\phi(i)}{n}\Big)$. We choose $m_\square$ such that $\psi_\square \simeq \psi_{m_\square} = \Theta(m_\square\sqrt{w_\square})$, that is $m_\square = \Theta(\delta^{-\frac{2}{3}}n^{-\frac{1}{3}}\log^{\frac{1}{3}} n)$. We can rewrite $\delta^{-\frac{2}{3}}n^{-\frac{1}{3}} = \delta^{-\frac{2}{5}}n^{\frac{1}{5}}(\delta n^2)^{-\frac{4}{15}}$. Finally, using $\delta > n^{-2}$, we have $m_\square = \tilde{O}(\delta^{-\frac{2}{5}}n^{\frac{1}{5}})$.

-2- Let $N_\sharp = \{(W_i, C_i)\}_{1\le i\le m_\sharp}$ be another family of nested pairs. Now, we define an angle $\psi_\sharp = \Theta(\delta^{-\frac{1}{5}}n^{-\frac{2}{5}})$. We set $w_{m_\square+i} = w_\sharp = \delta^{\frac{4}{5}}\psi_\sharp^{\frac{2}{5}}n^{-\frac{2}{5}}\log^{\frac{2}{5}} n = \delta^{\frac{18}{25}}n^{-\frac{14}{25}}\log^{\frac{2}{5}} n$ and $\phi(m_\square+i) = \phi(m_\square)+i\cdot n\sqrt{w_\sharp}$. We choose $m_\sharp$ such that $\psi_\sharp \simeq \psi_{m_\square+m_\sharp}$ The sampling condition gives $m_\sharp = \Theta(\psi_\sharp w_\sharp^{-\frac{1}{2}}) = \Theta(\delta^{-\frac{14}{25}}n^{-\frac{3}{25}}\log^{-\frac{1}{5}} n)$. Using $\delta > n^{-2}$, we can rewrite $\delta^{-\frac{14}{25}}n^{-\frac{3}{25}} = \delta^{-\frac{2}{5}}n^{\frac{1}{5}}(\delta n^2)^{-\frac{4}{25}}$. Finally, we have $m_\sharp = \tilde{O}(\delta^{-\frac{2}{5}}n^{\frac{1}{5}})$.

-3- Let $N_\diamond = \{(W_i, C_i)\}_{1\le i\le m_\diamond}$ be a third family of nested pairs. For $1 \le i \le m_\diamond$, we define $w_{m_\square+m_\sharp+i} = w_\diamond = \Theta(\delta^{\frac{4}{5}}n^{-\frac{2}{5}}\log^{\frac{2}{5}} n)$ and $\phi(m_\square + m_\sharp + i) = \phi(m_\square+m_\sharp)+i\cdot n\sqrt{w_\diamond}$. We choose $m_\diamond$ such that $\frac{\pi}{4} \simeq \psi_{m_\square+m_\sharp+m_\diamond}$. The sampling condition gives $m_\diamond = \Theta(\frac{\pi}{4}w_\diamond^{-\frac{1}{2}}) = \Theta(\delta^{-\frac{2}{5}}n^{\frac{1}{5}}\log^{-\frac{1}{5}} n)$.

Proof of Theorem 7. The family $N = \{(W_i, C_i)\}_{1\le i\le m=m_\square+m_\sharp+m_\diamond}$, augmented by its 8 symmetric copies for other octants, satisfies the hypotheses of Lemma 1.

Since the angle subtending any arc $W_i\cap\Gamma$ is $\Theta(w_\star^{\frac{1}{2}})$ where $\star\in\{\square,\sharp,\diamond\}$, an easy generalization of Lemma 3 ensures that each square $\mathfrak{P}(x)$ intersects $O(1)$ witnesses of that sub-family. It also ensures that $N_\square$ controls all directions between 0 and $\psi_\square$, $N_\sharp$ controls all direction between $\psi_\square$ and $\psi_\sharp$, and $N_\diamond$ controls all directions between $\psi_\sharp$ and $\frac{\pi}{4}$.

- Using Lemma 9, we have $|W_i\cap\mathfrak{P}(X)| \ge \log n$ and $|C_i\cap\mathfrak{P}(X)| = \Theta(\log n)$ for $1\le i\le m_\square$.

- Using Lemma 8 and adjusting the constant in the $\Theta(\cdot)$ defining $\psi_\sharp$, we have $|W_i \cap \mathfrak{P}(X)| = \Theta\Big(nw_\sharp^{\frac{5}{2}}\delta^{-2}\psi_i^{-1}\Big) = \psi_\sharp\psi_i^{-1}\log n \geq \log n$ and $|C_i \cap \mathfrak{P}(X)| = \Theta(\psi_\sharp\psi_i^{-1}\log n)$ for $m_\square \leq i \leq m_\square + m_\sharp$.
- Using Lemma 8 and adjusting the constant in the $\Theta(\cdot)$ defining $w_\diamond$, we have $|W_i \cap \mathfrak{P}(X)| = \Theta\Big(nw_\diamond^{\frac{5}{2}}\delta^{-2}\psi_i^{-1}\Big) = \frac{\pi}{4}\psi_i^{-1}\log n \geq \log n$ and $|C_i \cap \mathfrak{P}(X)| = \Theta(\psi_i^{-1}\log n)$ for $m_\square + m_\sharp \leq i \leq m_\square + m_\sharp + m_\diamond$.

We are now ready to apply Lemma 1:

$$\Omega\left(m - \sum_{1\leq i\leq m} e^{-|W_i\cap\mathfrak{P}(X)|}\right) \;\leq\; \mathbf{E}[\mathrm{ex}(\mathfrak{P}(X))] \;\leq\; n\sum_{1\leq i\leq m} e^{-|W_i\cap\mathfrak{P}(X)|} + \sum_{1\leq i\leq m}|C_i\cap\mathfrak{P}(X)|.$$

Since for all i, $|W_i \cap \mathfrak{P}(X)| \geq \log n$, it follows that the lower bound is $\Omega(m)$ and the first term of the upper bound is $O\big(nm\frac{1}{n}\big) = O(m)$. The second term evaluates to

$$\begin{aligned}
&\sum_{1\leq i\leq m} |C_i \cap \mathfrak{P}(X)| \\
&\quad\leq \sum_{1\leq i\leq m_\square} |C_i \cap \mathfrak{P}(X)| + \sum_{m_\square < i \leq m_\square + m_\sharp} |C_i \cap \mathfrak{P}(X)| + \sum_{m_\square + m_\sharp < i \leq m_\square + m_\sharp + m_\diamond} |C_i \cap \mathfrak{P}(X)| \\
&\quad\leq m_\square\Theta(\log n) + m_\sharp\Theta(\log n)\left(\sum_{1<i\leq m_\sharp}\psi_\sharp\psi^{-1}_{m_\square+i}\right) + m_\diamond\Theta(\log n)\left(\sum_{1<i\leq m_\diamond}\psi^{-1}_{m_\square+m_\sharp+i}\right) \\
&\quad\leq m\Theta(\log n)\left(\sum_{1\leq i\leq m_\sharp}\frac{1}{i} + \sum_{1\leq i\leq m_\diamond}\frac{1}{i}\right) \\
&\quad\leq m\Theta(\log n)(\log m_\sharp + \log m_\diamond)
\end{aligned}$$

Thus the expected size of the number of extreme points is $\tilde{\Theta}(m) = \tilde{\Theta}(\delta^{-\frac{2}{5}}n^{\frac{1}{5}})$ since $m = m_\square + m_\sharp + m_\diamond = \tilde{O}(\delta^{-\frac{2}{5}}n^{\frac{1}{5}}) + \tilde{O}(\delta^{-\frac{2}{5}}n^{\frac{1}{5}}) + \tilde{\Theta}(\delta^{-\frac{2}{5}}n^{\frac{1}{5}})$. □

5 Experimental results

In this section, we compare the theoretical bounds of Theorems 2 and 7 to noisy data obtained by two mechanisms: simulations of uniform noise using pseudo-random number generators and snap-rounding. Our implementation is based on the CGAL library [4]. Points are given in double precision floating-point arithmetic and convex hulls are computed using exact predicates.

Noise simulation. We first simulated a uniform noise of amplitude δ on a set X of n points in 2 and 3 dimensions, and counted the number of extreme points of the perturbed set Y. We consider $n = 10^i$ for $i = 3$ to 7 and $\delta = 10^j$ for $j = -7$ to 5. Our count of the number of extreme points is averaged over 1000 trials for $n \leq 10^6$ and over 100 trials for $n = 10^7$. In the plane, we applied uniform L^∞ and L^2 noise to the vertices of a near-regular n-gon; in 3D, we

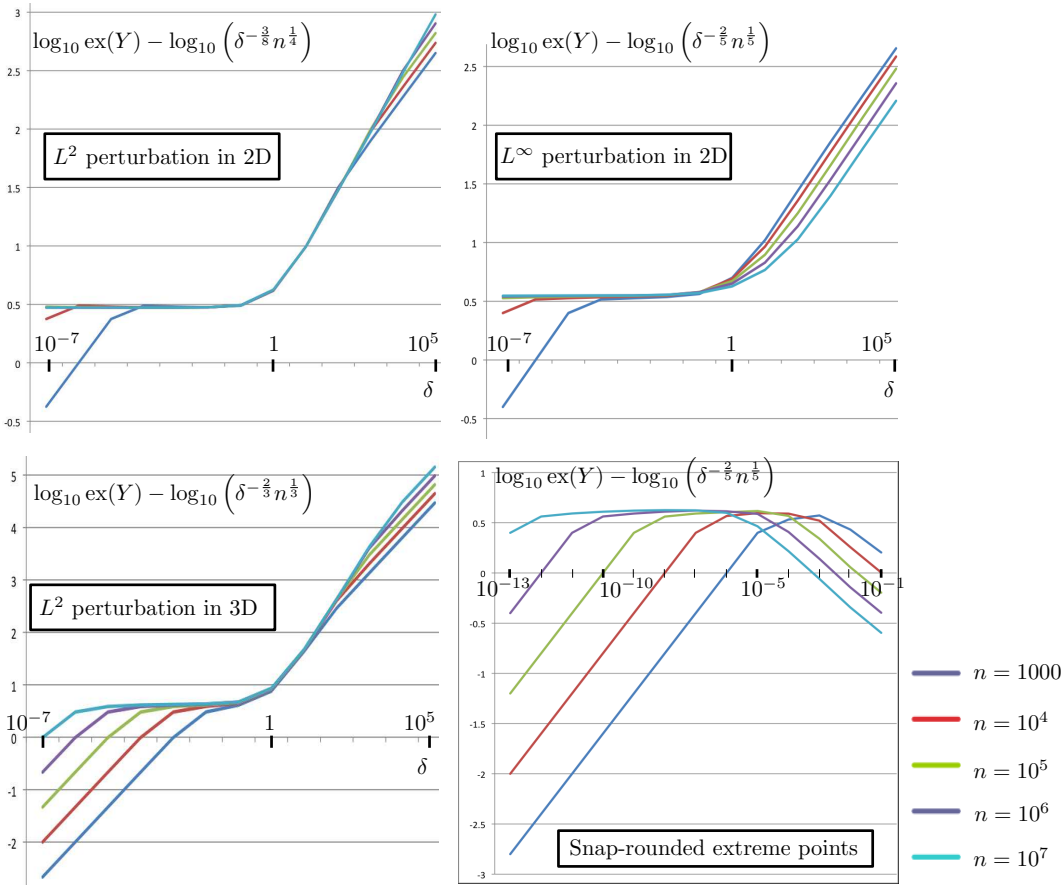


Figure 6: Ratios between practical and theoretical behaviors.

applied uniform L^2 noise to a random sample of the unit sphere. The data is presented in Appendix A.1–A.3. We plot in Figure 6 the ratio between the measure and our theoretical bound for $\mathbf{E}[\mathrm{ex}(\mathfrak{P}(X))]$ as a function of δ, using a logarithmic scale for both the horizontal and vertical axis. Each curve represents a different value of n.

We observe that all curves present a significant plateau, indicating that the behavior of the number of extreme points is correctly estimated by our formulas. More precisely, in the range $10^3 \leq n \leq 10^7$ these behaviors are modeled within a margin of error of 5% by the Table below.

Perturbation	L^2 in $\mathbb{R}^2$	L^∞ in $\mathbb{R}^2$	L^2 in $\mathbb{R}^3$
Estimation	$3\delta^{-\frac{3}{8}} n^{\frac{1}{4}}$	$4.75\delta^{-\frac{2}{5}} n^{\frac{1}{5}}$	$4.2\delta^{-\frac{2}{3}} n^{\frac{1}{3}}$
Range of δ	From $10/n^2$ to 0.1	From $1000/n^2$ to 0.1	From $1000/n$ to 0.01

Snap rounding. To measure the effect of snap-rounding, we started with a set X obtained by snap-rounding a regular n-gon inscribed in the unit circle of $\mathbb{R}^2$ to the non-regular grid of double precision float numbers. We ensured that in this first round of snap-rounding, all points in X remained extreme for all values of n considered; let us stress that in single precision float numbers, this is already *not* the case for $n = 10^5$. We then rounded X to a coarser regular grid of step δ, obtaining a set Y. Finally, we counted the number of extreme points of Y. We considered $n = 10^i$ for $i = 3, \dots, 7$ and $\delta = 10^j$ for $j = -13, \dots, 0$. The data is presented in Appendix A.4. The last plot in Figure 6 represents the ratio between the number of extreme points after snap-rounding to a grid of step δ and our theoretical bound for uniform L^∞ noise of amplitude δ (the predictions are better than those obtained using the bound for uniform L^2 noise). We use a logarithmic scale for the axes, and each curve represents a different value of n. We observe experimentally that for $\delta \geq \frac{1}{\sqrt{n}}$ the convex hull of the rounded point set is a discrete circle, whose size, $\Theta(\delta^{\frac{1}{3}})$, does not depend on n anymore, and for $\delta < \frac{10}{n^2}$, all points remain extreme once rounded. In-between, the behavior is predicted within a margin of error of 5% for $\delta \in \left[\frac{1000}{n^2}, \frac{1}{100\sqrt{n}}\right]$ by $4\delta^{-\frac{2}{5}} n^{\frac{1}{5}}$, that is our bound for uniform L^∞ noise.

6 Conclusion and future works

Theorems 2 and 7 generalize easily to $\left(\Theta\left(n^{\frac{1}{1-d}}\right), \Theta(1)\right)$-samples of smooth compact convex rotund surfaces. Extending our technique to non-uniform distributions with small support essentially requires to prove an analogue of Lemma 4; for distributions with non-compact support, Condition (b) in Lemma 1 should be replaced by an estimation of how many collectors may contain a given perturbed point.

It seems natural that the expected number of extreme points among n points distributed uniformly in a L^p ball reflects the behavior of the expected number of extreme points under uniform L^p perturbation for large values of δ; this would, however, indicate that the expected number of extreme points increases between $\delta = \Theta(1)$ and $\delta \to \infty$, after decreasing between $\delta = 0$ and $\delta = \Theta(1)$. Extending our result to a wider range of δ thus seems an interesting question.

The smoothed complexity model considers the maximum over an input X of the maximum expected complexity over small perturbations of X. In the light of Theorem 2, our restriction that X be in convex position keeps us away from the worst-case for certain values of δ: when uniform L^2 noise is applied to n points in the plane all with the same position, the expected number of extreme points is $\Theta(n^{\frac{1}{3}})$, which is larger than $\tilde{\Theta}\left(\delta^{-3/8} n^{1/4}\right)$ for $\delta = \Omega(n^{-2/9+\epsilon})$. What types of initial configuration determines the smoothed number of extreme points of a n-point set? Are they even the same for the uniform L^2 and L^∞ perturbations? To the best of our knowledge, both questions are open.

Acknowledgment

This work was initiated during the 8^{th} McGill - INRIA Workshop on Computational Geometry at the Bellairs institute. The authors wish to thank all the participants for creating a pleasant and stimulating atmosphere, in particular Nina Amenta and Sariel Har-Peled for useful discussions at the early stage of this work.

References

[1] D. Attali, J. D. Boissonnat, and A. Lieutier. Complexity of the Delaunay triangulation of points on surfaces the smooth case. In *Proc. 19th Annu. ACM Sympos. Comput. Geom.*, pages 201–210, 2003.

[2] J.-D. Boissonnat, D. Cohen-Steiner, B. Mourrain, G. Rote, and G. Vegter. Meshing of surfaces. In Jean-Daniel Boissonnat and Monique Teillaud, editors, *Effective Computational Geometry for Curves and Surfaces*, pages 181–229. Springer-Verlag, Mathematics and Visualization, 2006.

[3] F. Cazals and J. Giesen. Delaunay triangulation based surface reconstruction. In Jean-Daniel Boissonnat and Monique Teillaud, editors, *Effective Computational Geometry for Curves and Surfaces*, pages 231–276. Springer-Verlag, Mathematics and Visualization, 2006.

[4] CGAL Editorial Board. *CGAL User and Reference Manual*, 3.5 edition, 2009. http://www.cgal.org/Manual/3.5/doc_html/cgal_manual/packages.html.

[5] V. Damerow and C. Sohler. Extreme points under random noise. In *Proc. 12th Annu. European Sympos. Algorithms*, pages 264–274, 2004.

[6] Mark de Berg, Herman Haverkort, and Constantinos P. Tsirogiannis. Visibility maps of realistic terrains have linear smoothed complexity. In *SCG '09: Proceedings of the 25th annual symposium on Computational geometry*, pages 163–168, New York, NY, USA, 2009. ACM.

[7] Mark de Berg, Matthew Katz, A. Frank van der Stappen, and Jules Vleugels. Realistic input models for geometric algorithms. In *SCG '97: Proceedings of the thirteenth annual symposium on Computational geometry*, pages 294–303, New York, NY, USA, 1997. ACM.

[8] O. Devillers, J. Erickson, and X. Goaoc. Empty-ellipse graphs. In *Proc. 19th ACM-SIAM Sympos. Discrete Algorithms*, pages 1249–1257, 2008.

[9] J. Erickson. Nice point sets can have nasty Delaunay triangulations. *Discrete and Computational Geometry*, 30(1):109–132, 2003.

[10] J. Erickson. Dense point sets have sparse Delaunay triangulations. *Discrete and Computational Geometry*, 33:83–115, 2005.

[11] M Golin. How many maxima can there be? *Computational Geometry: Theory and Applications*, 2:335–353, 1993.

[12] M Golin and Na H.-S. On the average complexity of 3d-Voronoi diagrams of random points on convex polytopes. *Computational Geometry: Theory and Applications*, 25:197–231, 2003.

[13] M. J. Golin and H. S. Na. The probabilistic complexity of the Voronoi diagram of points on a polyhedron. In *Proc. 18th Annu. ACM Sympos. Comput. Geom.*, pages 209–216, 2002.

[14] J. E. Goodman, R. Pollack, and B. Sturmfels. Coordinate representation of order types requires exponential storage. In *STOC '89: Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 405–410, New York, NY, USA, 1989. ACM.

[15] S. Har-Peled. On the expected complexity of random convex hulls. http://valis.cs.uiuc.edu/~sariel/research/papers/notes/rand_hull.html, 1997.

[16] H. Raynaud. Sur l'enveloppe convex des nuages de points aleatoires dans R^n. *J. Appl. Probab.*, 7:35–48, 1970.

[17] A. Rényi and R. Sulanke. Über die konvexe Hülle von n zufällig gerwählten Punkten I. *Z. Wahrsch. Verw. Gebiete*, 2:75–84, 1963.

[18] A. Rényi and R. Sulanke. Über die konvexe Hülle von n zufällig gerwählten Punkten I. *Z. Wahrsch. Verw. Gebiete*, 3:138–147, 1964.

[19] D.A. Spielman and S.-H. Teng. Smoothed analysis: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51:385 – 463, 2004.

[20] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis: an attempt to explain the behavior of algorithms in practice. *Commun. ACM*, 52(10):76–84, 2009.

A Experimental results

This appendix presents the full experimental results that have been used for the curves presented at Section 5.

Tables below present experimental data used for the curves in Section 5. For each point, we indicate the average size of the convex hull and the standard deviation for 1000 experiments (100 experiments for 10M points).

A.1 Perturbation by a disk

Each vertex of the regular n-gon of radius 1 is perturbed uniformly in a disk of radius δ. Numbers in brown deviate by less than 5% and numbers in **bold brown** by less than 1% from the value $3\delta^{-\frac{3}{8}}n^{\frac{1}{4}}$. The 5% exactness is achieved for $\delta \in \left[\frac{10}{n^2}, 0.1\right]$. The 1% exactness is achieved for $\delta \in \left[\frac{100}{n^2}, 0.01\right]$.

2D Convex hull size
disk perturbation

δ	n				
	1 000	10 000	100 000	1 000 000	10 000 000
0.000 000 1	1 000 0	10 000 0	**22 671** 74	**39 631** 100	**70 420.5** 121
0.000 001	1 000 0	5 485 33	**9 441** 47	**16 692** 62	**29 703** 92
0.000 01	1 000 0	**2 267** 22	**3 963** 29	**7 040** 41	**12 531** 56
0.000 1	547.7 10	**943.8** 15	**1 670** 19	**2 969** 26	**5 281** 33
0.001	**226.7** 7.6	**396.7** 9.9	**703** 12	**1 253** 17	**2 229** 24
0.01	**94.5** 4.8	**167.5** 6.4	**298** 8.3	**531** 11	**943** 13
0.1	41.1 3.2	73.6 4.4	131 5.6	233 7.4	415 8.9
1	23.2 2.4	41.6 3.2	74.6 4.1	133 5.7	237 6.7
10	23.3 2.3	41.4 3.2	73.8 4.2	131 5.8	232 7.0
100	31.6 2.7	54.3 3.5	95.0 4.8	168 6.2	301 9.3
1 000	33.5 2.9	71.4 4.1	132 5.6	226 7.4	396 9.8
10 000	33.6 2.8	72.7 4.4	156 6.1	317 8.6	542 12
100 000	33.4 2.9	72.7 4.1	156 6.0	337 9.2	716 12

A.2 Perturbation by a square

Each vertex of the regular n-gon of radius 1 is perturbed uniformly in a square of side length δ. Numbers in brown deviate by less than 5% and numbers in **bold brown** by less than 2% from the value $4.75\delta^{-\frac{2}{5}}n^{\frac{1}{5}}$. The 5% exactness is achieved for $\delta \in \left[\frac{1000}{n^2}, 0.1\right]$.

2D Convex hull size
square perturbation

δ	n				
	1 000	10 000	100 000	1 000 000	10 000 000
0.000 000 2	1000 0	9994 2.3	21 245 73.5	34 739 95.3	55 839 107.8
0.000 002	1000 0	5202 32.7	8 613 45.1	13 955 58.1	**22 292** 73.5
0.000 02	999.4 0.73	2123 21.4	3 474 29.1	5 584 37.9	**8 896** 45.5
0.000 2	520.2 10.4	861.5 14.3	1 395 18.6	**2 231** 22.8	**3 544** 28.6
0.002	212.2 7.4	347.6 9.1	558.3 11.5	**890.4** 14.8	**1 411** 17.5
0.02	86.5 4.5	140.2 6.1	**225.1** 7.53	**358.2** 9.3	**569.6** 11.3
0.2	36.5 3.0	59.7 3.7	95.1 5.0	150.4 5.9	234.7 7.2
2	19.9 2.1	31.0 2.9	47.0 3.6	71.0 4.6	106.5 5.3
20	16.6 2.4	23.2 2.8	31.3 3.3	42.6 3.7	58.5 4.1
200	17.4 2.5	22.7 3.1	28.1 3.2	34.5 3.8	42.4 3.7
2 000	18.0 2.6	23.5 3.1	28.7 3.5	34.3 3.9	39.7 4.0
20 000	18.1 2.6	24.0 3.3	29.7 3.7	35.0 3.9	40.4 4.7
200 000	17.9 2.7	24.2 3.1	30.2 3.8	35.9 4.0	40.4 4.5

A.3 Perturbation by a sphere

Each point of a random sampling of the 3D unit sphere is perturbed in a ball of radius δ. Numbers in brown deviate by less than 10% and numbers in **bold brown** by less than 3% from the value $4.2\delta^{-\frac{2}{3}}n^{\frac{1}{3}}$. The 10% exactness is achieved for $\delta \in \left[\frac{100}{n}, 0.1\right]$. The 3% exactness is achieved for $\delta \in \left[\frac{1000}{n}, 0.01\right]$.

3D Convex hull size ball perturbation					
δ	n				
	1 000	10 000	100 000	1 000 000	10 000 000
0.000 000 1	1 000	10 000	99 999	999 781	9 842 030
0.000 001	1 000	9 999.9	99 977	984 203	6 493 270
0.000 01	999.9	9 997	98 420	649 353	1 784 770
0.000 1	999	9 842	64 935	178 489	**410 642**
0.001	984	6 495	17 864	**41 084**	**90 852**
0.01	651	1 797	**4 141**	**9 159**	**19 975**
0.1	**190**	445	991	2 168	4 684
1	75.5	174	388	849	1 838
10	93.0	210	464	1 010	2 186
100	134	402	927	2 009	4 333
1 000	135	441	1 396	4 061	9 266
10 000	136	442	1 410	4 467	14 009
100 000	136	442	1 410	4 474	14 146

A.4 Snap rounding

The regular n-gon centered at the origin and with radius 1 is rounded on the grid of step δ. For $\delta \geq \frac{1}{\sqrt{n}}$, we get a discrete circle (the number of extreme points is independent of n). For $\delta < \frac{10}{n^2}$, all points stay extreme during rounding. In between, we can model the number of extreme points using the formula for the square perturbation. The 5% exactness is achieved for $\delta \in \left[\frac{1000}{n^2}, \frac{1}{100\sqrt{n}}\right]$. Numbers in brown deviate by less than 5% from the value $4\delta^{-\frac{2}{5}} n^{\frac{1}{5}}$.

2D Convex hull size
snap rounding

δ	n				
	1 000	10 000	100 000	1 000 000	10 000 000
10^{-13}	1 000	10 000	100 000	1 000 000	10 000 000
10^{-12}	1 000	10 000	100 000	1 000 000	5 790 000
10^{-11}	1 000	10 000	100 000	1 000 000	2 470 000
10^{-10}	1 000	10 000	100 000	578 824	1 020 000
10^{-9}	1 000	10 000	100 000	246 519	417 700
10^{-8}	1 000	10 000	57 764	102 526	168 140
10^{-7}	1 000	10 000	24 653	41 828	66 639
10^{-6}	1 000	5 855	10 118	16 370	25 049
10^{-5}	1 000	2 487	41 43	6 145	7 327
10^{-4}	540	980	1 474	1 617	1 649
0.001	236	332	348	348	348
0.01	68	72	72	72	72
0.1	16	16	16	16	16
1	4	4	4	4	4

Centre de recherche INRIA Sophia Antipolis – Méditerranée
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399